P3-18

Real-Time Free Viewpoint Rendering System for Face-to-face Video Conference

Chia-Ho Pan, Sheng-Chieh Huang[1], Yu-Lin Chang, Chung-Jr Lian, and Liang-Gee Chen, *Fellow, IEEE*
DSP/IC Design Lab, Graduate Institute of Electronics Engineering
and Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan
[1]*Department of Electrical and Control Engineering, National Chiao-Tung University, HsinChu, Taiwan*

Abstract **– Face-to-face communication experience plays an important role in videoconference applications. In this paper, we propose view angle rectification and view synthesis methods to generate the preferred view at observer's side. The proposed free viewpoint rendering system achieves face-to-face communication experience for real-time videoconference application.**

I. Introduction

Videoconference is developed by the researchers for a long time. As the new devices and new technologies of multimedia coding and transmission are improved, the relative products of this application have increased rapidly. Face-to-face communication is the first and essential step to create a convincing impression for this kind of application. But the lack of the natural of personal communication affects the growth of videoconference product.

Most existed systems use a fixed camera location to capture the images at speaker's side. These video conferencing equipments does not support natural mutual gaze experience because of the angular disparity between the camera mounted on top of the monitor and the location of the facial image on the screen. When the speaker looks at the observer's eyes on the screen, the observer gets the impression that the speaker is looking down. This is due to the displacement between camera and the facial image. Only when the speaker looks at the camera, it appears to the observer that they are looking each other. But in this case, the speaker can not know whether the observer is looking at him or not. This cause the unnatural using experience in most existed systems for videoconference applications.

Several works have been proposed to overcome these limitations [1]-[2]. Most previous works focus on the angular deviation when the conference window is located at a predefined location on the screen. The line of vision shown in conference window is rectified according to the relationship between the location of the predefined conference window and the camera. But it is rectified regardless of the conference window location at observer's side in conventional videoconference systems. However, the angular deviation of the face-to-face communication depends on the relationship between the camera location, the observer's eyes at speaker's side, and the location of conference window on the screen at observer's side.

For current videoconference application, the user opens several windows for using videoconference and other events at the same time. Fig. 1 shows the emerging using behavior of desktop video conferencing application. The angular derivation from the speaker's side is displayed at the observer's side as a θ angle. Even if this angular derivation is considered, the line of vision of speaker is still changed because of the uncertain location of conference window at observer's side as shown in the right part of Fig. 1. To solve this problem, we address this angular derivation issue at observer's side in this paper. We proposed a real-time free viewpoint rendering system to generate the preferred view at observer's side. The synthesized view rectifies the face position and eye direction and conveys the natural image of the changes in gaze and facial expression for face-to-face communication.

This work was partially supported by ASUSTek Computer Inc..

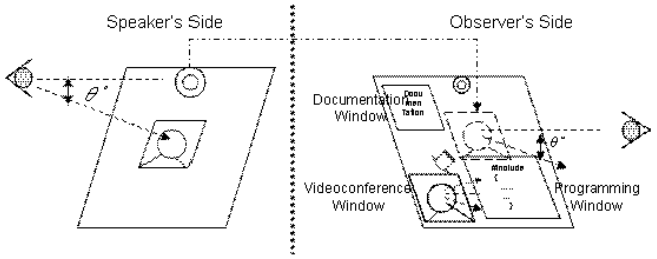

Fig.1. Scenario of desktop videoconference system.

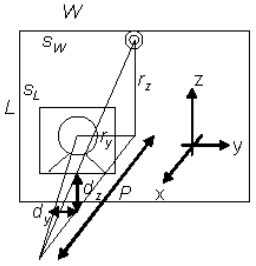

Fig.2. View angle rectification of observer's side.

II. Free Viewpoint Rendering System

The proposed free viewpoint rendering system can be overviewed as two parts: view angle rectification and view synthesis of preferred view for the observer's side.

A. View angle rectification

In the previous works, they usually assume the conference window at the observer's side is in the predefined location on the screen. When the conference window at observer's side is not in the predefined location on the screen, it has another angular deviation between the conference window and predefined location as shown in Fig. 1. This display angular

1-4244-1459-8/08/$25.00 ©2008 IEEE

deviation at the observer's side is also important aspect to affect the perceptual impression. In Fig.2, we rectify this view angle between the camera and the conference window of the preferred location on the screen. In order to capture an applicable size of videoconference, we assume the distance P = 50 cm from the screen to the observer in desktop video conferencing application. The displacement of display angular deviation can be derived from the mouse location of the speaker eyes in the conference window on the screen. Then the rectification angle at observer's side is

$$d_y = \tan^{-1}\left(\frac{r_y}{s_w} * W / P\right), \qquad d_z = \tan^{-1}\left(\frac{r_z}{s_L} * L / P\right), \quad (1)$$

where d_y, d_z are the deviation angle at observer's side that should be rectified, s_W, s_L are the resolution settings of the screen, r_y, r_z are the displacements of window resolution between capture camera and the conference window, and W, L are the width and length of the monitor.

B. *View synthesis*

After getting the rectification angle, we have to generate preferred view of the conference window on the screen. We use depth-based scheme to generate the preferred view because it is more nature than those schemes without depth information. The depth map generation and depth image based rendering (DIBR) are used to synthesize the preferred view for the observer. The depth information is generated according to the concept of depth from motion method. The depth from motion method analyzes the motion detection of object motion and its time difference from previous frame and current frame. The depth map is generated by some modification of our block-based depth estimation [3] and then smoothes the depth information after depth estimation. Although the depth information is not sensitive than the image of conference window [4], the resolution of depth map can be generated as smaller resolution. In our case, the resolution of depth map is half in width and height of the original image size.

With the original conference image and the depth map, we can generate the preferred view for the observer according to the view angle rectification. The depth map and the pixel information are mapping to the vertex of 3D model. For the real-time issue, the DIBR is accelerated by the graphic processor unit (GPU) on the graphic card. By doing this, 3D Graphics transform and color space conversion can be accelerated. We use the OpenGL2.0 to call the GPU for the view synthesis of preferred view.

III. Simulations

In this simulation, the images of videoconference are captured from the webcam mounted on the top of the monitor. Fig. 3(a) is the captured image from the webcam. Fig. 3(b) shows the synthesized view in the middle on the screen as the predefined location in previous works. Fig. 3(c) and (d) are the synthesized images when the conference window are located at the left-up and right of the monitor with angular rectification d_y = -15.535°, d_z = -8.866°, and d_y = 6.277°, d_z = -11.309°, respectively. Any conference window location can be synthesized at observer's side. From the simulated images, the face position and eye direction of the synthesized images are rectified and the proposed system achieves the natural image for face-to-face communication which is the most important aspect in videoconference applications.

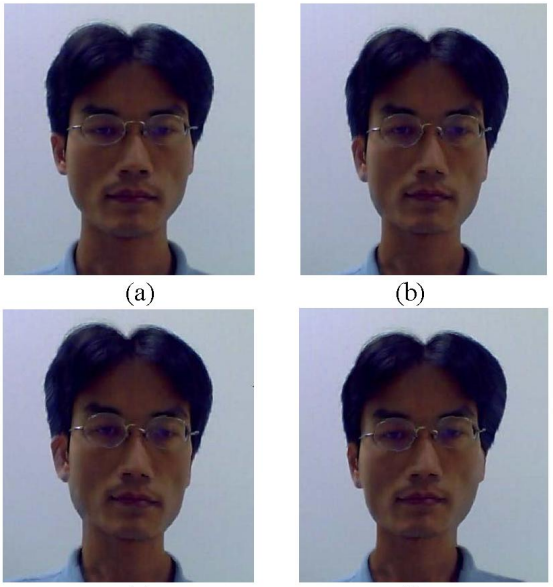

Fig.3. (a) The captured image from the camera mounted on the top of the screen of speaker's side; (b), (c), (d) are the images with the watching window placed at the middle, left-up, and right on the monitor of observer's side

IV. Conclusions

In this paper, we propose a free viewpoint rendering system with view angle rectification and view synthesis for the preferred view at observer's side in desktop video conferencing application. In addition, we examined how well it can establish face-to-face communication experience. From the simulations, we have proved that the proposed scheme can sufficiently convey nonverbal information such as changes in gaze and facial expression of natural human-centered communication at observer's side. Moreover, it is effective to synthesize the preferred view at any location of conference window on the screen and achieves the natural image with face-to-face communication in videoconference application.

References

[1] Gemmell, J., Zitnick, C.L., Kang, T., Toyama, K., and Seitz, S., "Gaze-awareness for videoconferencing: a software approach." IEEE Trans. on Multimedia, pp. 26-35, 2000.

[2] R. Yang, and Z. Zhang, "Gaze-awareness for videoconferencing : a software approach." IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 26, no. 7, pp. 956-920, Jun. 2004.

[3] Y.-M. Tsai, Y.-L. Chang, and L.-G. Chen, "Symmetric Trinocular Dense Disparity Estimation For Car Surrounding Camera Array, " in SPIE Symposium on Electronics Imaging, Visual Communications and Image Processing, 2007.

[4] P. Harman, J. Flack, S. Fox, and M. Dowley, "Rapid 2d to 3d conversion," Proc. SPIE Stereoscopic Displays and Virtual Reality Systems IX, vol. 4660, pp. 78–86, May 2002.